FIG. 1A

2736 bp

```
   1 ATGGCGCCGA GCTCCCGTC ACCCGCCGCG CCTACACGCG TCTCTGGGCG
  51 GAAGCGCGCC GCCAAGGCCG AGGAGATCCA CCAGAACAAG GAGGAGGAGG
 101 AGGAGGTCGC GGCGGCGTCG TCCGCCAAGC GCAGCCGCAA GGCGGCATTT
 151 TCCGGGAAGA AGCCCAAGTC GCCCCCCAAG AAGCCAAGC CGGGGAGGAA
 201 GAAGAAGGGG GATGCCGAGA TGAAGGAGCC CGTGGAGGAC GACGTGTCCG
 251 CCGAGGAGCC CGACGAGGAG GAGTTGGCCA TGGGCGAGGA GGAGGCCGAG
 301 GAGCAGGCCA TGCAGGAGGA GGTGGTTGCG GTCGCGGCGG GGTCACCCGG
 351 GAAGAAGAGG GTGGGGAGAA GGAACGCCGC CGCCGCCGCT GGCGACCACG
 401 AGCCGGAGTT CATCGGCAGC CCTGTTGCCG CGGACGAGGC GCGCGCAAC
 451 TGGCCCAAGC GCTACGGCCG CAGCACTGCC GCAAAGAAAC CGGATGAGGA
 501 GGAAGAGCTC AAGGCCAGAT GTCACTACCG GAGCGCTAAG GTGGACAACG
 551 TCGTCTACTG CCTGGGGGAT GAGGTCTATG TCAAGGCTGG AGAAAACGAG
 601 GCAGATTACA TTGGCCGCAT TACTGAATTT TTTGAGGGGA CTGACCAGTG
 651 TCACTATTTT ACTTGCCGTT GGTTCTTCCG AGCAGAGGAC ACGGTTATCA
 701 ATTCTTTGGT GTCCATAAGT GTGGATGGCC ACAAGCATGA CCCTAGACGT
 751 GTTTTTCTTT CTGAGGAAAA GAACGACAAT GTGCTTGATT GCATTATCTC
 801 CAAGGTCAAG ATAGTCCATG TTGATCCAAA TATGGATCCA AAAGCCAAGG
 851 CTCAGCTGAT AGAGAGTTGC GACCTATACT ATGACATGTC TTACTCTGTT
 901 GCATATTCTA CATTTGCTAA TATCTCGTCT GAAAATGGGC AGTCAGGCAG
 951 TGATACCGCT TCGGGTATTT CTTCTGATGA TGTGGATCTG GAGACGTCAT
1001 CTAGTATGCC AACGAGGACA GCAACCCTTC TTGATCTGTA TTCTGGCTGT
1051 GGGGGCATGT CTACTGGTCT TTGCTTGGGT GCAGCTCTTT CTGGCTTGAA
1101 ACTTGAAACT CGATGGGCTG TTGATTTCAA CAGTTTTGCG TGCCAAGTTT
1151 TAAAATATAA TCATCCACAG ACTGAGGTGC GAAATGAGAA AGCCGATGAG
1201 TTTCTTGCCC TCCTTAAGGA ATGGGGCAGT CTATGCAAAA AATATGTCCA
1251 AGATGTGGAT TCAAATTTAG CAAGCTCAGA GGATCAAGCG GATGAAGACA
1301 GCCCTCTTGA CAAGGACGAA TTTGTTGTAG AGAAGCTTGT CGGGATATGT
1351 TATGGTGGCA GTGACAGGGA AAATGGCATC TATTTTAAGG TCCAGTGGGA
1401 AGGATACGGC CCTGAGGAGG ATACATGGGA ACCGATTGAT AACTTGAGTG
1451 ACTGCCCGCA GAAAATTAGA GAATTTGTAC AAGAAGGGCA CAAAAGAAAG
1501 ATTCTCCCAC TGCCTGGTGA TGTTGATGTC ATTTGTGGAG GCCCACCATG
1551 CCAAGGTATC AGTGGGTTTA ATCGGTACAG AAACCGTGAT GAGCCACTCA
1601 AAGATGAGAA AAACAAACAA ATGGTGACTT TCATGGATAT TGTGGCGTAC
1651 TTGAAGCCCA AGTATGTTCT CATGGAAAAT GTGGTGGACA TACTCAAATT
1701 TGCGGATGGT TACCTAGGAA AATATGCTTT GAGCTGCCTT GTTGCTATGA
1751 AGTACCAAGC GCGGCTTGGA ATGATGGTGG CTGGTTGCTA TGGTCTGCCA
1801 CAGTTCAGGA TGCGTGTGTT CCTCTGGGGT GCTCTTTCTT CCATGGTGCT
1851 CCCTAAGTAT CCTCTGCCCA CCTATGATGT TGTAGTACGT GGAGGAGCCC
1901 CTAATGCCTT TTGGCAATGT ATGGTTGCAT ATGACGAGAC ACAAAAACCA
1951 TCCGTGAAAA AAGCCTTGCT TCTTGGCGAT GCAATTTCAG ATTTACCAAA
2001 GGTTCAAAAT CACCAGCCTA ACGATGTGAT GGAGTATGGT GGTTCCCCCA
2051 AGACCGAATT CCAGCGCTAC ATTCGACTCA GTCGTAAAGA CATGTTGGAT
2101 TGGTCCTTCG GTGAGGGGGC TGGTCCAGAT GAAGGCAAGC TCTTGGATCA
2151 CCAGCCTTTA CGGCTTAACA ACGATGATTA TGAGCGGGTT CAACAGATTC
2201 CTGTCAAGAA GGGAGCCAAC TTCCGCGACC TAAAGGGCGT GAGGGTTGGA
2251 GCAAACAATA TTGTTGAGTG GGATCCAGAA ATCGAGCGTG TGAAACTTTC
2301 ATCTGGGAAA CCACTGGTTC CTGACTATGC AATGTCATTC ATCAAGGGCA
2351 AATCACTCAA GCCGTTTGGG CGCCTGTGGT GGGACGAGAC AGTTCCTACA
2401 GTTGTAACCA GAGCAGAGCC TCACAACCAG GTTATAATTC ATCCGACTCA
2451 AGCAAGGGTC GTCACTATCC GGGAGAACGC AAGGTTACAG GGTTCCCCG
2501 ATTACTACCG ATTGTTTGGC CCGATCAAGG AGAAGTACAT TCAAGTCGGG
2551 AACGCAGTGG CTGTCGTTGT TGCCCGGGCA CTGGGCTACT GTGTGGGGCA
2601 AGCCTACCTG GGTGAATCTG AGGGGAGTGA CCCTCTGTAC CAGCTGCCTC
2651 CAAGTTTCAC CTTTGTTGGA GGACGCACTG CGGGGCAGGC GAGGGCTCT
2701 CCTGTTGGCA TCCCTGCAGG GGAGGTAGTT GAGCAG
```

FIG. 1B

```
   1  AGAGCAGCAG CAGCTACCGC AGCCCCTGCC ATGGCGCCGA GCTCCCCGTC
  51  ACCCGCCGCG CCTACACGCG TCTCTGGGCG GAAGCGCGCC GCCAAGGCCG
 101  AGGAGATCCA CCAGAACAAG GAGGAGGAGG AGGAGGTCGC GGCGGCGTCC
 151  TCCGCCAAGC GCAGCCGCAA GGCGGCATCT TCCGGGAAGA AGCCCAAGTC
 201  GCCCCCCAAG CAGGCCAAGC CGGGGAGGAA GAAGAAGGGG GATGCCGAGA
 251  TGAAGGAGCC CGTGGAGGAC GACGTGTGCG CCGAGGAGCC CGACGAGGAG
 301  GAGTTGGCCA TGGGCCAGGA GGAGGCCGAG GAGCAGGCCA TGCAGGAGGA
 351  GGTGGTTGCG GTCGCGGCGG GGTCACCCGG GAAGAAGAGG GTGGGGAGAA
 401  GGAACGCCGC CGCCGCCGCT GGCGACCACG AGCCGGAGTT CATCGGCAGC
 451  CCTGTTGCCG CGGACGAGGC GCGCAGCAAC TGGCCCAAGC GCTACGGCCG
 501  CAGCACTGCC GCAAAGAAAC CGGATGAGGA GGAAGAGCTC AAGGCCAGAT
 551  GTCACTACCG GAGCGCTAAG GTGGACAACG TCGTCTACTG CCTCGGGGAT
 601  GACGTCTATG TCAAGGCTGG AGAAAACGAG GCAGATTACA TTGGCCGCAT
 651  TACTGAATTT TTTGAGGGGA CTGACCAGTG TCACTATTTT ACTTGCCGTT
 701  GGTTCTTCCG AGCAGAGGAC ACGGTTATCA ATTCTTTGGT GTCCATAAGT
 751  GTGGATGGCC ACAAGCATGA CCCTAGACGT GTTTTTCTTT CTGAGGAAAA
 801  GAACGACAAT GTGCTTGATT GCATTATCTC CAAGGTCAAG ATAGTCCATG
 851  TTGATCCAAA TATGGATCCA AAAGCCAAGG CTCAGCTGAT AGAGAGTTGC
 901  GACCTATACT ATGACATGTC TTACTCTGTT GCATATTCTA CATTTGCTAA
 951  TATCTCGTCT GAAAATGGGC AGTCAGGCAG TGATACCGCT TCGGGTATTT
1001  CTTCTGATGA TGTGGATCTG GAGACGTCAT CTAGTATGCC AACGAGGACA
1051  GCAACCCTTC TTGATCTGTA TTCTGGCTGT GGGGGCATGT CTACTGGTCT
1101  TTGCTTGGGT GCAGCTCTTT CTGGCTTGAA ACTTGAAACT CGATGGGCTG
1151  TTGATTTCAA CAGTTTTGCG TGCCAAAGTT TAAAATATAA TCATCCACAG
1201  ACTGAGGTGC GAAATGAGAA AGCCGATGAG TTTCTTGCCC TCCTTAAGGA
1251  ATGGGCAGTT CTATGCAAAA AATATGTCCA AGATGTGGAT TCAAATTTAG
1301  CAAGCTCAGA GGATCAAGCG GATGAAGACA GCGCTCTTGA CAAGGACGAA
1351  TTTGTTGTAG AGAAGCTTGT CGGGATATCT TATGGTGGCA GTGACAGGGA
1401  AAATGGCATC TATTTTAAGG TCCAGTGGGA AGGATACGGC CCTGAGGAGG
1451  ATACATGGGA ACCGATTGAT AACTTGAGTG ACTCCCGCA GAAAATTAGA
1501  GAATTTGTAC AAGAAGGGCA CAAAAGAAAG ATTCTCCCAC TGCCTGGTGA
1551  TGTTGATGTC ATTTGTGGAG GCCCACCATG CCAAGGTATC AGTGGGTTTA
1601  ATCGGTACAG AAACCGTGAT GAGCCACTCA AAGATGAGAA AAACAAACAA
1651  ATGGTGACTT TCATGGATAT TGTGGCGTAC TTGAAGCCCA AGTATGTTCT
1701  CATGGAAAAT GTGGTGGACA TACTCAAATT TGCGGATGGT TACCTAGGAA
1751  AATATGCTTT GAGCTGCCTT GTTGCTATGA AGTACCAAGC GCGGCTTGGA
1801  ATGATGGTGG CTGGTTGCTA TGGTCTGCCA CAGTTCAGGA TGCGTGTGTT
1851  CCTCTGGGGT GCTCTTTCTT CCATGGTGCT CCCTAAGTAT CCTCTGCCCA
1901  CCTATGATGT TGTAGTACGT GGAGGAGCCC CTAATGCCTT TTCGCAATGT
1951  ATGGTTGCAT ATGACGAGAC ACAAAAACCA TCCCTGAAAA AAGCCTTGCT
2001  TCTTGGCGAT GCAATTTCAG ATTTACCAAA GGTTCAAAAT CACCAGCCTA
2051  ACGATGTGAT GGAGTATGGT GGTTCCCCCA AGACCGAATT CCAGCGCTAC
2101  ATTCGACTCA GTCGTAAAGA CATGTTGGAT TGGTCCTTCG GTGAGGGGGC
2151  TGGTCCAGAT GAAGGCAAGC TCTTGGATCA CCAGCCTTTA CGGCTTAACA
2201  ACGATGATTA TGAGCGGGTT CAACAGATTC CTGTCAAGAA GGGAGCCAAC
2251  TTCCGCGACC TAAAGGGCGT GAGGGTTGGA GCAAACAATA TTGTTGAGTG
2301  GGATCCAGAA ATCGAGCGTG TGAAACTTTC ATCTGGGAAA CCACTGGTTC
2351  CTGACTATGC AATGTCATTC ATCAAGGGCA AATCACTCAA GCCGTTTGGA
2401  CGCCTGTGGT GGGACGAGAC AGTTCCTACA GTTGTAACCA GAGCAGAGCC
2451  TCACAACCAG GTTATAATTC ATCCGACTCA AGCAAGGGTC CTCACTATCC
2501  GGGAGAACGC AAGGTTACAG GGCTTCCCCG ATTACTACCG ATTGTTTGGC
2551  CCGATCAAGG AGAAGTACAT TCAAGTCGGG AACGCAGTGG CTGTCCCTGT
2601  TGCCCGGGCA CTGGGCTACT GTCTGGTGCA AGCCTACCTG GGTGAATCTG
2651  AGGGGAGTGA CCCTCTGTAC CAGCTGCCTC CAAGTTTCAC CTCTGTTGGA
2701  GGACGCACTG CGGGGCAGGC GAGGGCCTCT CCTGTTGGCA CCCCTGCAGG
2751  GGAGGTAGTT GAGCAGTAAA AGGATGACAG ATCTGAGCTG AGCTGG
```

FIG. 2A 912 amino acids

```
   1  MAPSSPSPAA PTRVSGRKRA AKAEEIMQNK EEEEEVAAAS SAKRSRRAAS
  51  SGKKPKSPPK QAKPGRKKKG DAEMKEPVED DVCAEEPDEE ELAMGEEEAE
 101  EQAMQEEVVA VAAGSPGKKR VGRRNAAAAA GDHEPEFIGS PVAADEARSN
 151  WPKRYGRSTA AKKPDEEEEL KARCHYRSAK VDNVVYCLGD DVYVKAGENE
 201  ADYIGRITEF FEGTDQCHYF TCRWFFRAED TVINSLVSIS VDGHKHDPRR
 251  VFLSEEKNDN VLDCIISKVK IVHVDPNMDP KAKAQLIESC DLYYDMSYSV
 301  AYSTFANISS ENGQSGSDTA SGISSDDVDL ETSSSMPTRT ATLLDLYSGC
 351  GGMSTGLCLG AALSGLKLET RWAVDFNSFA CQSLKYNHPQ TEVRNEKADE
 401  FLALLKEWAV LCKKYVQDVD SNLASSEDQA DEDSPLDKDE FVVEKLVGIC
 451  YGGSDRENGI YFKVQWEGYG PEEDTWEPID NLSDCPQKIR EFVQEGHKRK
 501  ILPLPGDVDV ICGGPPCQGI SGFNRYRNRD EPLKDEKNKQ MVTFMDIVAY
 551  LKPKYVLMEN VVDILKFADG YLGKYALSCL VAMKYQARLG MMVAGCYGLP
 601  QFRMRVFLWG ALSSMVLPKY PLPTYDVVVR GGAPNAFSQC MVAYDETQKF
 651  SLKKALLLGD AISDLPKVQN HQPNDVMEYG GSPKTEFQRY IRLSRKDMLD
 701  WSFGEGAGPD EGKLLDHQPL RLNNDDYERV QQIPVKKGAN FRDLKGVRVG
 751  ANNIVEWDPE IERVKLSSGK PLVPDYAMSF IKGKSLKPFG RLWWDETVPT
 801  VVTRAEPHNQ VIIHPTQARV LTIRENARLQ GFPDYYRLFG PIKEKYIQVG
 851  NAVAVPVARA LGYCLGQAYL GESEGSDPLY QLPPSFTSVG GRTAGQARAS
 901  PVGTPAGEVV EQ
```

FIG. 2B

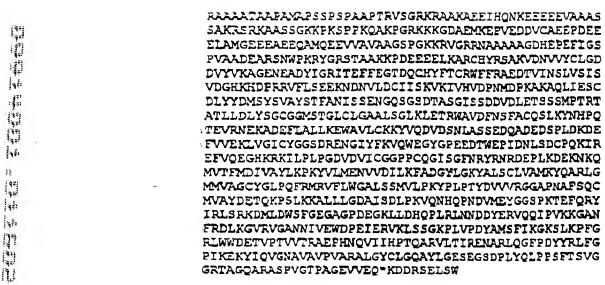

RAAAATAAPANGLPSSPSPAAPTRVSGRKRAAKAEEIHQNKEEEEEVAAAS
SAKRSRKAASSGKKPKSPPKQAKPGRKKKGDAEMKEPVEDDVCAEEPDEE
ELAMGEEEAEEQAMQEEVVAVAAGSPGKKRVGRRNAAAAAGDHEDEFIGS
PVAADEARSNWPKRYGRSTAAKKPDEEEELKARCHYRSAKVDNVVYCLGD
DVYVKAGENEADYIGRITEFFEGTDQCHYFTCRWFFRAEDTVINSLVSIS
VDGHKHDPRKVFLSEEKNDNVLDCIISKVKIVHVDPNMDPKAKAQLIESC
DLYYDMSYSVAYSTFANISSENGQSGSDTASGISSDDVDLETSSSMPTRT
ATLLDLYSGCGGMSTGLCLGAALSGLKLETRWAVDFNSFACQSLKYNHPQ
TEVRNEKADEFLALLKEWAVLCKKYVQDVDSNLASSEDQADEDSPLDKDE
FVVEKLVGICYGGSDRENGIYFKVQWEGYGPEEDTWEPIDNLSDCPQKIR
EFVQEGHKKRILPLPGDVDVICGGFPCQGISGFNRYRNRDEPLKDEKNKQ
MVTFMDIVAYLKPKYVLMENVVDILKFADGYLGKYALSCLVAMKYQARLG
MMVAGCYGLPQFRMRVFLWGALSSMVLPKYPLPTYDVVVRGGAPNAFSQC
MVAYDETQKPSLKKALLLGDAISDLPKVQNHQPNDVMEYGGSPKTEFQRY
IRLSRKDMLDWSFGEGAGPDEGKLLDHQPLRLNNDDYERVQQIPVKKGAN
FRDLKGVRVGANNIVEWDPEIERVKLSSGKPLVPDYAMSFIKGKSLKPFG
RLWWDETVPTVVTRAEPHNQVIIHPTQARVLTIRENARLQGFPDYYRLFG
PIKEKYIQVGNAVAVPVARALGYCLGQAYLGESEGSDPLYQLPPSFTSVG
GRTAGQARASPVGTPAGEVVEQ*KDDRSELSW

FIG. 3

Primer	Sequence 5' - 3'
1F	TGGTTGCTATGGTCTGCCACAGTTCAG
1R	CCAGCTCAGCTCAGATCTGTCATCCTTT
Seq2FN	CGAAAGCTAATCTACACAAACAGC
Seq2RN	GATCCTCTGAGCTTGCTAAATTTG
3R	CTCATCTTGGAGTGGCTCATCAC
S3F	GAGCACATGAGGGAGAGTGTTG
S3R	TCTCTAATTTTCTGCGGGCAG
4F	CCTCTGCCCACCTATGATGTTGTA
5F	TAAAGGGCGTGAGGGTTGGA
7F	TCACATTTGTCATGGCAGGTTATC
8eF	CTGAGGAAAAGAACGACAATGTGC
8eR	GCAATCAAGCACATTGTCGTTCTTTTCCTC
9eF	GAAGAAGAGGGTGGGGAGAAGGAACG
9eR	TTCTTTGCGGCAGTGCTGCG
11iF	GTATTGAATTGATTCTCAACTAGTGCAC
11iR	CAGGCTCAACGGCGATG
12iF	TATGCTTCATCACATAGACCCAAGTC
12iR	GATAGACCTAATGCCAAATGAGATTAAG
13iF	GCGATCTTCAGTCTCCACCATC
13iR	GAAGACGTGCCTCCATGTTTCATC
14F	GTTGGTTCTTCCGAGCAGAGG
14R	GACTGCCACATATCTTATTAATCGC
15F	GCATGTGTCAGCAATTGCTTACATTC
15R	CCTCTGCTCGGAAGAACCAAC
16F	CTGTTCGGAGATTCATGCATGATG
16R	GGAGAACAGAATGGTTGATTCAATGG
17F	GCACTTCACTCTCCTGGCAAACC
17R	CGGTACGCTGCTGCTGCTCTC
18F	CCATAGCATCTCACATATCGCAAGG
18R	GGAAAGAAGGCAGTTAGTTGTAAATGGG
MU	AGAGAAGCCAACGCCAWCGCCTCYATTTCGTC
RaceRT	CTACAACATCATAGTTGGGCAGAGG
AP2 marathon	ACTCACTATAGGGCTCGAGCGGC
T7	TAATACGACTCACTATAGGG
Sp6	GATTTAGGTGACACTATAG
M13F	GTTTTCCCAGTCACGAC
M13R	CAGGAAACAGCTATGAC

FIG. 4

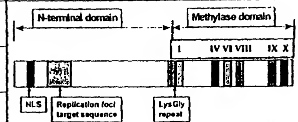
Gene Name	Organism	Function	General Structure
DNMT1/Dnmt1	human/mouse	maintenance	
MET1	*Arabidopsis*	maintenance	
Zmet1	maize	putative maintenance	
DNMT3/Dnmt3	human/mouse	*de novo*	
Zmet3	maize	putative *de novo*	
DRM	*Arabidopsis*	putative *de novo*	
CMT1	*Arabidopsis*	undetermined (putative CpNpG)	
Zmet2a	maize	CpNpG (maintenance and/or *de novo*)	

Figure 5

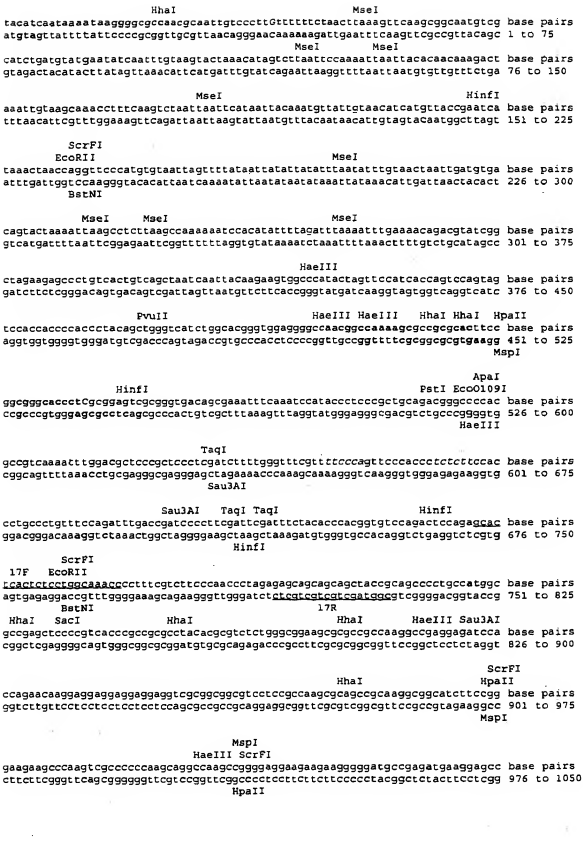

HhaI
MseI
base pairs
1 to 75
MseI
MseI
base pairs
76 to 150
MseI
HinfI
base pairs
151 to 225
ScrFI
EcoRII
MseI
base pairs
226 to 300
BstNI
MseI
MseI
MseI
base pairs
301 to 375
HaeIII
base pairs
376 to 450
PvuII
HaeIII HaeIII
HhaI HhaI
HpaII
base pairs
451 to 525
MspI
ApaI
HinfI
PstI Eco0109I
base pairs
526 to 600
HaeIII
TaqI
base pairs
601 to 675
Sau3AI
Sau3AI
TaqI TaqI
HinfI
base pairs
676 to 750
HinfI
ScrFI
17F
EcoRII
base pairs
751 to 825
BstNI
17R
HhaI
SacI
HhaI
HhaI
HaeIII Sau3AI
base pairs
826 to 900
ScrFI
HhaI
HpaII
base pairs
901 to 975
MspI
MspI
HaeIII ScrFI
base pairs
976 to 1050
HpaII

FIG. 5

Continued

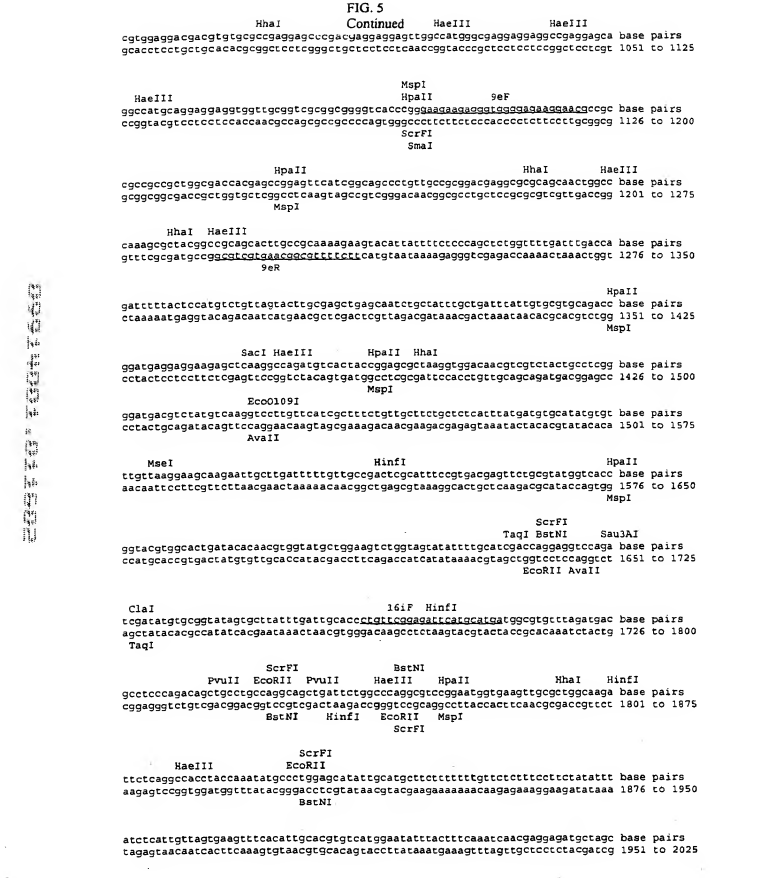


FIG. 5 Continued

```
                                        EcoRV
attgaggtgtgtgataattattacatactagaagatatcgtgcatgttgccattgggattgcgaagaatgtggaa base pairs
taacttccacacactattaataatgtatgatcttctatagcacgtacaacggtaaccctaacgcttcttacacctt 2026 to 2100

                          MseI    MseI
atgatgttggtagcttgtattaagaggttaacagttagtgggatgacatgactattagttagagatgatgtggat base pairs
tactacaaccatcgaacataatttctccaattgtcaatcaccctactgtactgataatcaatcttctactacacca 2101 to 2175

agtaagtgggatatgatgtagatgacttgtgtgttgagacagaactataacatggagttggaaatgggagcagca base pairs
tcattcaccctatactacatctactgaacacacaactctgtcttgatattgtacctcaacctttaccctcgtcgt 2176 to 2250

                                                          MseI
tggtcaaacataccctaaatgcctgtcctcacacaatgtggtgattggtgtatagtctggtgttaaaagctggat base pairs
accagtttgtatgggatttacggacagagatgtgttacaccactaaccacatatcagaccacaattttcgacctа 2251 to 2325

                                                                 ScrFI
     HinfI                          MseI       XbaI                EcoRII
actttgattctgttgaagattgtcacaccctgaatttaaggacaaatctagatacatcttcatatgtgcaccaggat base pairs
tgaaactaagacaacttctaacagtgtgggcttaaattcctgtttagatctatgtagagtatacacgtggtccta 2326 to 2400
                                                                 BstNI

agtgtatagataccaatgtcataatctttattacacgacgataatgtcttacaaaatatctggtgttacaagatg base pairs
tcacatatctatggttacagtattagaaaataatgtgctgctattacagaatgttttatagaccacaatgttctac 2401 to 2475

                     MseI                MseI
caccttttcaacatgtttaatgccgcaaactgttttaattaaacagaatgcagtgttttgaacaaaaaaatgctgc base pairs
gtggaaagttgtacaaattacgacgttgacaaaattaatttgtcttacgtcacaaaacttgtttttttacgacg 2476 to 2550
                                    MseI

                              15iF                        Sau3AI    HinfI MseI
tttatcctgcatcttgttttgcatgtgtcagcaattgcttacatccattatgatctcgagattccttaaattt base pairs
aaataggacgtagaacaaaacgtacacagtcgttaacgaatgtaaggtaatactagagactctaagaaatttaaa 2551 to 2625

ctagcatgatgaaagtatttactaattcaactgaacacaaacattgtttgaatgaacaaggcaacacggatgctt base pairs
gatcgtactacttttcataaatgattaagttgactttgtgtttgtaacaaacttactgttccgttgtgcctacgaa 2626 to 2700

                                                          MseI
ggaataatggttgtgtataatatcacttagtggttttgctctcacaccacatctttcatgggtctttaataata base pairs
ccttattaccaacacatattatagtgaatcaccaaaacgagagtgtggtgtagaaagtacccaagaaattattat 2701 to 2775

          MseI                                                          HaeIII
gttactgactttaagtttcttattccttttgtctatcttagctggagaaaacgaggcagattacattggccgc base pairs
caatgactgaaattcaaagaataaggaaaaacagatagaatcgacctcttttgctccgtctaacgtaaccggcg 2776 to 2850

                                                    14eF
attactgaatttttgaggggactgaccagtgtcactatttacttgccgttggttcttccgagcagaggacacg base pairs
taatgacttaaaaaactcccctgactggtcacagtgataaaatgaacggcaaccaagaaaggctcgtctcctgtgc 2851 to 2925
                                                    15eR

gtgtgtatttagtatttgtcattctatgcatgtgtggatttttctggaatgtggaaaacatacagcactctctc base pairs
cacacatаaatcatаaaacagtaagatacgtacacacctaaaaagaccttacacctttgtatgtcgtgagagag 2926 to 3000

                                  MseI   HaeIII           HaeIII
tacaccacacacacttctagtatatgtgtacacgttaatgggccaacactagacacatggccccaacatcccccct base pairs
atgtggtgtgtgtgaagatcatatacacacgtgcgcaattacccggttgtgatctgtgtaccgggttgtaggggga 3001 to 3075

          EcoRV
caagatgggcgatagatatcaatcatcccccatcttgctacataacacatcacactcttttactcctatacccctta base pairs
```

FIG. 5
Continued

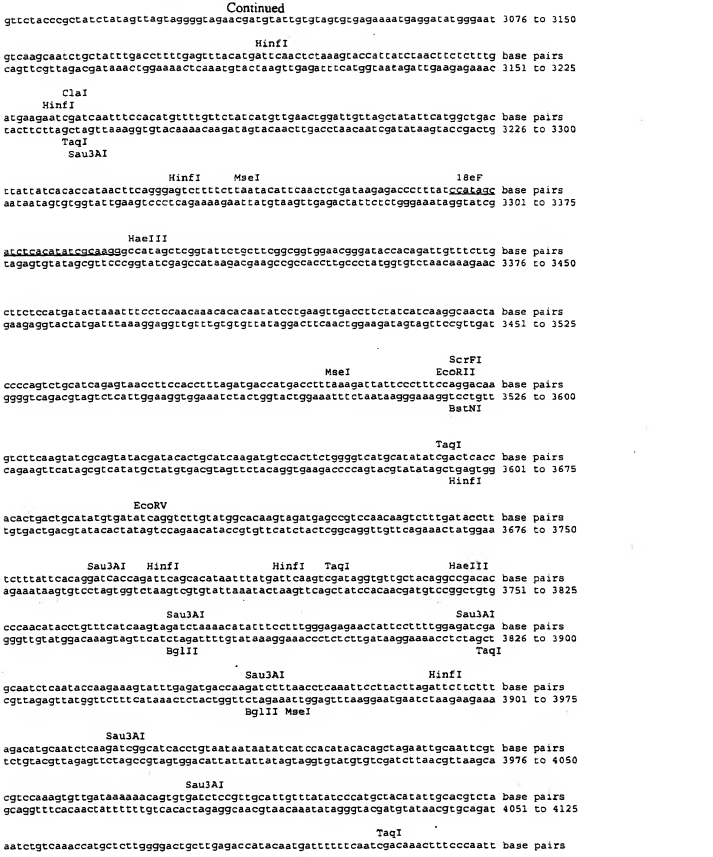
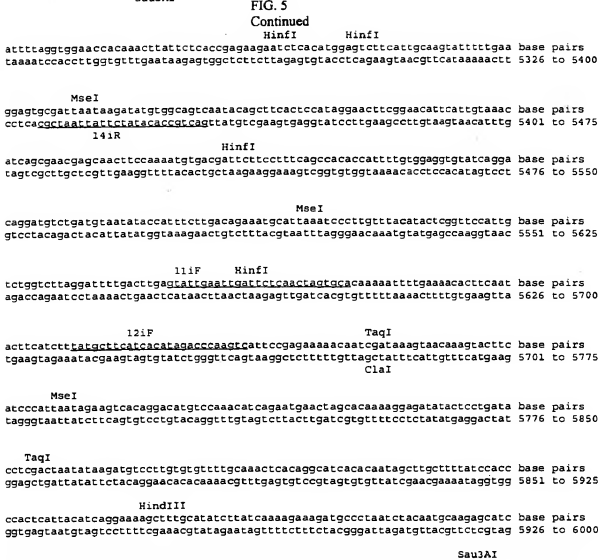


FIG. 5
Continued

```
ttagacagtttggtacgagaaccccctgacgaacctcggtatgttactaaaaaagttagctgtttgaaagggttaa 4126 to 4200

                    ScrFI
                  EcoRII  Sau3AI
gtctcaggctttgacaatccaggagggatctccatatagacctcctcttgcaaatcaccatgtaagaaagcattc base pairs
cagagtccgaaactgttaggtcctccctagaggtatatctggaggagaacgtttagtggtacattctttcgtaag 4201 to 4275
                    BstNI


 MseI                HaeIII                Sau3AI
ttaacatctagttgatacaagggccatccaaaatttgcagcacaagagatcaatgtccttacagtactcatttttt base pairs
aattgtagatcaactatgttcccggtaggttttaaacgtcgtgttcttctagttacaggaatgtcatgagtaaaaa 4276 to 4350


gccactggtgcaaatgtctcatcataatcaattccatatgtttgactataccctcttgcaaccaatcttgcttta base pairs
cggtgaccacgtttacagagtagtattagttaaggtatacaaactgatatgggagaacgttggttagaacgaaat 4351 to 4425


tatcgttctacccttccttctgggttttgcttcacagtgaataccctttacaactaactgccttctttccttta base pairs
atagcaagatgggaaggaagacccaaaacgaagtgtcacttatgggtaaatgttgattgacggaagaaaggaaat 4426 to 4500
                                                    18iR
                          XbaI      MseI
ggtagtttctcaaattcccaagtttgatttttttctagagctttaagctcctccaacattgcctcacgccagtta base pairs
ccatcaaagagtttaagggttcaaactaaaaaaagatctcgaaatccgaggaggttgtaacggagtgcggtcaat 4501 to 4575


gaattacattgtgcttctttccaatctcttggaattgctacggaatgcaatgatgcaacaaatgctctatatgat base pairs
cttaatgtaacacgaagaaaggttagagaaccttaacgatgccttacgttactacgttgtttacgagatatacta 4576 to 4650


                                                    HinfI
ggtgacaaagacgcatatgagacataattgctaatgtcatgttcatatccatacctttgttggggggactccagct base pairs
ccactgtttctgcgtatactctgtattaacgactacagtacaagtatagatatggaacaacccccctgaggtcga 4651 to 4725

          HhaI
ttagcacgcgctcctttcgtattgcaatgggcaaatcataagtgtcataatcttcagtttctccatgagacgtc base pairs
aatcgtgcgcgaggaaaagcataacgttacccgtttagtattcacagtattagaagtcaaagaggtactctgcag 4726 to 4800


aaaggtacatttatagcctctaatgtgtttggagagaactgctcagtacttgatgctgaattggttccaggagcc base pairs
tttccatgtaaatatcggagattacacaaacctctcctgacgagtcatgaactacgacttaaccaaagtcctcgg 4801 to 4875


tgaggttgcacatgggacttttcttctgtatatacttcgcccctatatcgtaagtcgtctccacaagatttatta base pairs
actccaacgtgtaccctgaaagaagaacatatatgaagcgggaatatagcattcagcagaggtgttctaaataat 4876 to 4950


ttctcgtgactaggatgtgctccaattccacttggcattacttgcatcttttgagaagcaccaatcaccacttcc base pairs
aagagcactgatcctacacagaggttaagtgaaccgcaatgaacgtagaaaactcttcgtggttagtggtgaagg 4951 to 5025


                        HinfI                       TaqI
attttatttggttgttgttccattgaatcaaccattctgttcccccctctcgactagcttcatctgtgctagta base pairs
taaaataaaccaacaacaaggttaacttagttggtaagacaagaggggagagctgatcgaagtagacacgatcat 5026 to 5100
                        16iR

       HinfI              Sau3AI
gagacagaatcaagaaaaaaatttagatctgtcttctcaccatagaaaggcacagtctctctaaatgtaacatcc base pairs
ctctgtcttagttcttttttttaaatctagacagaagagtggtatctttccgtgtcagagagatttacattgtagg 5101 to 5175
                        BglII


                          HinfI                             PstI
atgcttacaaacaaacgtcgttcactaggactcccaacacttgtatccctttgccctgcaggatatccaacaaaa base pairs
tacgaatgttgttgcagcaagtgatcctgaggttgtgaacatagggaaaacgggacgtcctatagggttgtttt 5176 to 5250
                                                          EcoRV

                    BamHI                       Sau3AI
atgcacttcacagcacgaggatccaacttccccacctgaggtctatgatcttgacaaaacatgtacatccaaaa base pairs
tacgtgaagtgtcgtgctcctaggttgaaggggtggactccagatactagagacctgttttgtacatgtaggtttt 5251 to 5325
```

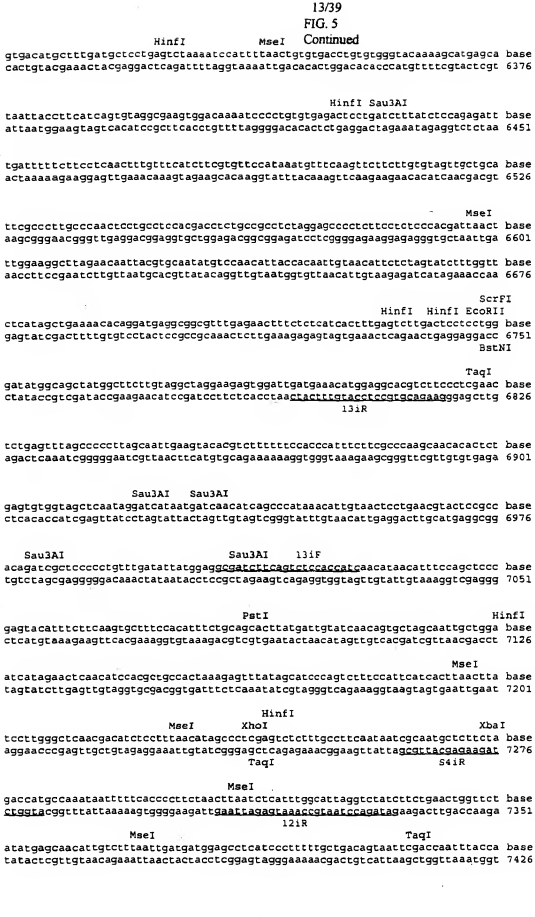
FIG. 5
Continued




```
attttaggtggaaccacaaacttattctcaccgagaagaatctcacatggagtctttcattgcaagtatttttgaa base pairs
taaaatccaccttggtgtttgaataagagtggctcttcttagagtgtacctcagaagtaacgttcataaaaactt 5326 to 5400

ggagtgcgattaataagatatgtggcagtcaatacagcttcactccataggaacttcggaacattcattgtaaac base pairs
cctcagcctaattattctatacaccgtcagttatgtcgaagtgaggtatccttgaagccttgtaagtaacatttg 5401 to 5475

atcagcgaacgagcaacttccaaaatgtgacgatttcttcctttcagccacaccatttttgtggaggtgtatcagga base pairs
tagtcgcttgctccgttgaaggttttacactgctaagaaggaaagtcggtgtggtaaaacacctccacatagtcct 5476 to 5550

caggatgtctgatgtaatataccatttcttgacagaaatgcattaaatccctttgtttacatactcggttccattg base pairs
gtcctacagactacattatatggtaaagaactgtctttacgtaatttagggaacaaatgtatgagccaaggtaac 5551 to 5625

tctggtcttaggattttgacttgagtattgaattgattctcaactagtgcacaaaaattttgaaaacacttcaat base pairs
agaccagaatcctaaaactgaactcataacttaactaagagttgatcacgtgtttttaaaactttgtgaagtta 5626 to 5700

acttcatcttgtagcttcatcacatagacccaagtcattccgagaaaaacaatcgataaagtaacaaagtacttc base pairs
tgaagtagaaatacgaagtagtgtatctgggttcagtaaggctctttttgttagctatttcattgtttcatgaag 5701 to 5775

atcccattaatagaagtcacaggacatgtccaaacatcagaatgaactagcacaaaaggagatatactcctgata base pairs
tagggtaattatcttcagtgtcctgtacaggtttgtagtcttacttgatcgtgttttccctctatatgaggactat 5776 to 5850

cctcgactaatataagatgtccttgtgtgttttgcaaactcacaggcatcacacaatagcttgcttttatccacc base pairs
ggagctgattatattctacaggaacacacaaaacgtttgagtgtccgtagtgtgttatcgaacgaaaataggtgg 5851 to 5925

ccactcattacatcaggaaaagctttgcatatcttatcaaaagaaagatgccctaatctacaatgcaagagcatc base pairs
ggtgagtaatgtagtccttttcgaaacgtatagaatagttttctttctacgggattagatgttacgttctccgtag 5926 to 6000

actgcaacctccttctcttccattcttgttgccagcatagtgcatattgtaccattagtcccctcatgatccata base pairs
tgacgttggaggaagagaaggtaagaacaacggtcgtatcacgtataacatggtaatcaggggagtactaggtat 6001 to 6075

taccacaatccattacgcctggtagctgtcccaagtctcttccctgtttccctctcctgaattaaacaattatct base pairs
atggtgttaggtaatgcggaccatcgacagggttcagagaagggacaaaggggagaggacttaatttgttaataga 6076 to 6150

cgatcaagaataatacgacaatccaattgatcaaccaaggcacttagtgatatcaagttgacaggaaaggttggc base pairs
gctagttcttattatgctgttaggttaactagttggttccgtgaatcactatagttcaactgtcctttccaaccg 6151 to 6225

acatacaaaactgatgacaacttaatagatggagtgcattgcactgtgccaacacccttgatgggttgtggtgta base pairs
tgtatgtttgactactgttgaattatctacctcacgtaacgtgacacggttgtgggaactacccaacaccacat 6226 to 6300

ccatcagcagtttgtataatttctttacgtgtgggggatatcttatatatgatgtaaatttcactggacgtgcct base pairs
ggtagtcgtcaaacatattaaagaaatgcacacccccctatagaatatatactacatttaagtgacctgcacgga 6301 to 6375
```

13/39

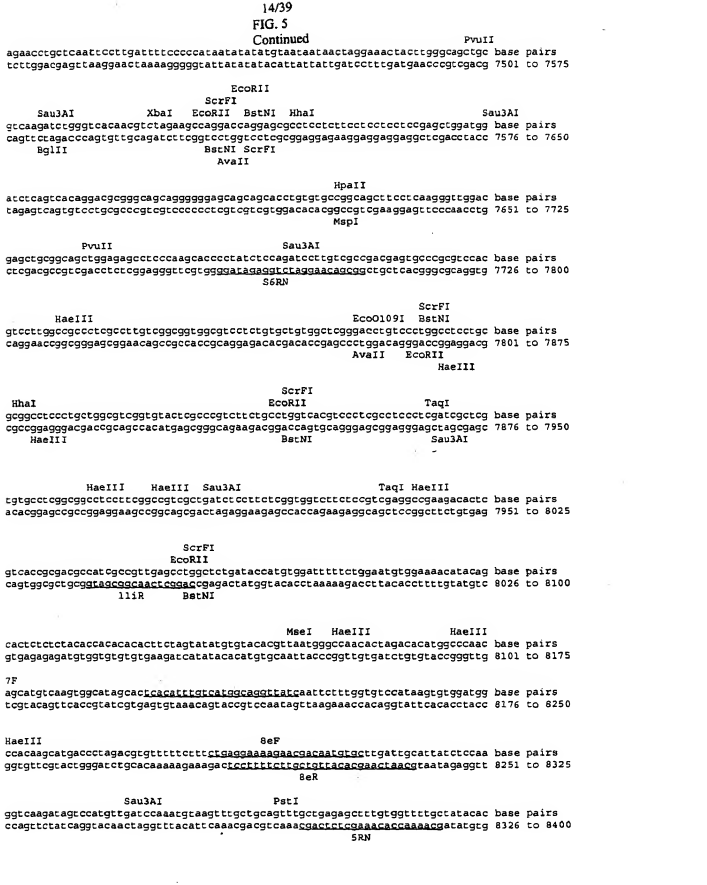
FIG. 5 Continued



14/39

FIG. 5

Continued



base pairs 7501 to 7575

base pairs 7576 to 7650

base pairs 7651 to 7725

base pairs 7726 to 7800

base pairs 7801 to 7875

base pairs 7876 to 7950

base pairs 7951 to 8025

base pairs 8026 to 8100

base pairs 8101 to 8175

base pairs 8176 to 8250

base pairs 8251 to 8325

base pairs 8326 to 8400

FIG. 5

Continued

```
                                        BamHI                    PvuII
ataatgttttctgactaccattgttttgtcgcctacttgccttagatggatcccaaaagccaaggcccagctgatag base pairs
tattacaaagactgatggtaacaaaacaacggatgaacggaatcctacctaggttttcggtttccgagtcgactatc 8401 to 8475
                                        Sau3AI

agagttgcgaccctatactatgacatgtcttactctgttgcatatttctacatttgctaatatctcgtctggtaatt base pairs
tctcaacgctggatatgatactgtacagaatgagacaacgtataagatgtaaacgatctatagagcagaccattaa 8476 to 8550

                                                    MseI
ccttctgcatcatctttttttggttgactagctgaatgcagttagctttgccaaagagttaaatacatgagttgtt base pairs
ggaagacgtagtagaaaaaaccaactgatcgacttacgtcaatcgaaacggtttctcaatttatgtactcaacaa 8551 to 8625

        TaqI                                                MseI    MseI
cctgcactcgaaaagggatgtcaataatgtccacaaacccctgaaaatgtatttttagatacttaacttgttaagt base pairs
ggacgtgagcttttccctacagttattacaggtgtttgagacttttacatataaaatctatgaattgaacaattca 8626 to 8700

cagtaaaacctgtcagatacttgggttttgggtacgattaccatccttatgtgagtaaaactcgtcaagggatgt base pairs
gtcattttggacagtctatgaacccaaaacccatgctaatggtaggaatacactcatttgagcagttccctaca 8701 to 8775

                                        TaqI       Seq2FN
caatgacgtgttgattgtgtattagatatctgtttgttcgaaagctaatctacacaaaacagcttatgtaatgta base pairs
gttacgcacaactaacacatatctcacaagacaaacaagctttcgattagatgtgtttgtcgaatacattacat 8776 to 8850

                            HindIII                                 HaeIII
aaacctcaaacaaacttgcctcttcataagcttaggtttataggattagcgtttagtgcatgtaaggcctatttg base pairs
tttggagcttgtcttgaacggagaagtattcgaatccaaatatcctaatcgcaaatcacgtacattccggataaac 8851 to 8925

                        BstNI           ScrFI                       ScrFI
        HaeIII        SacI            EcoRII                TaqI  EcoRII
cttcacggcctccctgccgagctcctggctagacagccatcctggccgtaggtgcccgaaatcgaacacctggga base pairs
gaagtgccggagggacggctcgaggaccgatctgtcggtaggaccggcatccacgggccttagcttgtggaccct 8926 to 9000
                        EcoRII          BstNI                       BstNI
                          ScrFI                 HaeIII

                          ScrFI
                        EcoRII
gccacgtttgcactagcaggttttcctgggtgcaaaccaaacacgcctatagtgttcaagtataactgaattggt base pairs
cggtgcaaacgtgatcgtccaaaaggacccacgttggtttgtgcggatatcacaagttcatattgacttaacca 9001 to 9075
                          BstNI

                      MseI
gctcaccttttgtctaatagcttaagttttggttttcatcggtgcatgcaactccatactcaatagtcaatatga base pairs
cgagtggaaacagattatcgaattcaaaaaccaaaagtagccacgtacgttgaggtatgagtatcagttatact 9076 to 9150

                                    BstNI
                        XhoI      HinfI
tatagtgttcaagcatagaactcctcgagtttgaatcctggcaggggcaatcaaataaaataattgcagcttaccc base pairs
atatcacaagttcgtatcttgagagctcaaacttaggaccgtccccgttagtttattttattaacgtcgaatggg 9151 to 9225
                        TaqI      EcoRII
                                    ScrFI

                        SfiF
ctatttctacgtttgagcacatgagggaagtgttgaattataagtgtgttctccatctttctctaacagatgaa base pairs
gataaagatgcaaactcgtgtactccctctcacaacttaatattcacacaagagggtagaaagagattgcctactt 9226 to 9300

                                HinfI           MseI                MseI
ctggtttgtgcatgtaactcaatatgatatttgagtcaaatgtttacttttaaaatcatagttgatgcaatttaat base pairs
gaccaaacacgtacattgagttatactataaactcagtttacaaatgaaatttagtatcaactacgttaaatta 9301 to 9375
```

16/39

FIG. 5

Continued

aacatatttttttggtctcgtgtgagggagtgtacgtataactgaactgcacacatttccttatagctttaggttt base pairs
ttgtataaaaaaaccagagcacactccctcacatgcatattgacttaacgtgtgtaaaggaatatcgaatccaaa 9376 to 9450

Sau3AI

ttgactgcaactgttggtgcatgtagctcaataactaaagttgatctggacagtctacagtgaataagtttgaca base pairs
aactgacgttgacaaccacgtacatcgagttattgatttgaactagacctgtcagatgtcacttattcaaactgc 9451 to 9525

cttgtaaaatgtgcatgtatttttacaaacgctggcactttttttcctaatagaaaatgggcagtcaggcagtgat base pairs
gaacattttacacgtacatataaaaatgtttgcgaccgtgaaaaaaggattatcttttacccgtcagtccgtcacta 9526 to 9600

Sau3AI

accgcttcgggtatttcttctgatgatgtggatctggagacgtcatctagtatgccaacgaggacagcaacccttt base pairs
tggcgaagcccataaagaagactactacacctagacctctgcagtagatcatacggttgctcctgtcgttgggaa 9601 to 9675

Sau3AI

cttgatctgtattctggctgtgggggcatgtctactggtctttgcttgggtgcagctctttctggcttgaaacttt base pairs
gaactagacataagaccgacacccccgtacagatgaccagaaacgaacccacgtcgagaaagaccgaacttttgaa 9676 to 9750

Sau3AI

gaaactgtaatcttctaactagtcatctgttggatagaatatgttccacgatcttcagaacttattctattgttctg base pairs
ctttgacattagaagattgatcagtagacaacctatcttatacaagtgctagagtcttgaataagataacaagac 9751 to 9825

MseI

gcttgcagcgatgggctgttgatttcaacagttttgcgtgccaaagtttaaaatataatcatccacagactgagg base pairs
cgaacgtcgctacccgacaactaaagttgtcaaaacgcacggtttcaaatttatattagtaggtgtctgactcc 9826 to 9900

HinfI

tatggatagtaaacttcatcttggattccatctgttctgtcagctactcttacaaagtgtctggatttttggatg base pairs
atacctatcatttgaagtagaacctaaggtagacaagacagtcgatgagaatgtttcacagacctaaaaacctac 9901 to 9975

MseI

taggtgccgaaatgagaaagccgatgagtttcttgccctccttaaggaatgggcagttctatgcaaaaaatatgtc base pairs
atccacgctttactctttcggctactcaaagaacgggaggaattccttacccgtcaagatacgttttttatacag 9976 to 10050

HinfI Sau3AI

caagatgtggattcaaatttagcaagctcagaggatcaagcggatgaagacagccctcttgacaaggacgaattt base pairs
gttctacacctaagtttaaatcgttcgaagtctcctagttcgcctacttctgtcgggagaactgttcctgcttaaa 10051 to 10125
Seq2RN

HindIII MseI

gttgtagagaagcttgtcgggatatgttatggtggcagtgacagggaaaatggcatctattttaaggtacttcag base pairs
caacatctcttcgaacagccctatacaataccaccgtcactgtccctttaccgtagataaaattccatgaagtc 10126 to 10200

HinfI MseI

tgtcatttgttcatttctacttgattccaacaaaaaaatcaattacttaagcctgtcaaacgatggatatttctg base pairs
acagtaaacaagtaaagatgaactaaggttgttttttttagttaatgaattcggacagtttgctcacctataaagac 10201 to 10275

PstI HaeIII

tatattttgctgtaacgctagatttctgcaggtccagtgggaaggatacggccctgaggaggatacatgggaacc base pairs
atataaaacgacattgcgatctaaagacgtccaggtcacccttcctatgccgggactcctcctatgtacccttgg 10276 to 10350
AvaII

gattgataacttgaggttagtgtatggtatatcgtctgcttgttgccttgtatacctatttgcatctaatccttg base pairs
ctaacttattgaactccaatcacatacctatagccagacgaacaacgggaacatatggataaacgtagattaggaac 10351 to 10425
Seq1RN

HinfI

FIG. 5
Continued

```
ttttgcaaaccagtgactgcccgcagaaaattagagaatttgtacaagaagggcacaaaagaaagattctcccac base pairs
aaaacgtttggtcactgacgggcgtctttttaatctctttaaacatgttctttcccgtgttttcttctaagagggtg 10426 to 10500
                  S3eR

                                                                    EcoRV
tgcctgtgagtatttagttcgttgtgattttgctcgccatttgtttagctccccctttttattcggtgatatctg base pairs
acggacactcataaatcaagcaacactaaaacgagcgataaacaaatcgaggggaaaaaataaaccactatagac 10501 to 10575

                                                 HaeIII                 MseI
cctattttattcttttcaaagggtgatgttgatgtcatttgtggaggcccaccatgccaaggtatcagtgggttta base pairs
ggataaaaataagaaagtttcccactacaactacagtaaacacctccgggtggtacggttccatagtcacccaaat 10576 to 10650

          3F
atcggtacagaaaccgtgatgagccactcaaagatgagaaaaacaaacaaatggtgactttcatggatattgtgg base pairs
tagccatgtctttggcactactcggtgagtttctactcttttttgtttgtttaccactgaaagtacctataacacc 10651 to 10725
                3R

cgtacttgaagcccaagtatgttctcatggaaaatgtggtggacatactcaaatttgcggatggttacctaggaa base pairs
gcatgaacttcgggttcatacaagagtaccttttcacaccacctgtatgagttttaaacgcccaccaatggatcctt 10726 to 10800

                                                 HhaI                   1F
aatatgctttgagctgccttgttgctatgaagtaccaagcgcggcttggaatgatggtggcrggttgctatggtc base pairs
ttatacgaaacttcgacggaacaacgatacttcatggttcgcgccgaaccttactaccaccgaccaacgatacag 10801 to 10875

                                Race2A
taccacagttcaggatgcgtgctgtaccttcgggtgctctttctcccctgtctgttctgtacctgccgttttta base pairs
acggtgtcaagtcctacgcacacatggagacccccacgagaaaagaaggtccagacaagacatggaacgacaaaat 10876 to 10950
                                Race2B

                                                      ScrFI
                HinfI                               EcoRII
tatgcttcgctagattcatattgcactgttggctgctggctaaccaggtgtacgtgtatttgacaatttaggtgc base pairs
atacgaagcgatctaagtataacgtgacaaccgacgaccgattggtccacatgcacatcaaactgttaaatccacg 10951 to 11025
                                                      BstNI

  Race1A          4F
tccctaagcatcctctgcccaacttatgatgttgtagtacgtggaggagcccctaatgcctttcggtgagtgcaat base pairs
agggattcatcggagacgggttgatactacaacatcatgcacctcctcggggattacggaaagccactcacgtta 11026 to 11100
  Race1B          RaceRT

cacaaaccactactatgaaatcatgtggaatgtgtaaaatacgctgaccaactgaatttgttgcagcaatgtatg base pairs
gtgtttggtgatgatactttagtacaccttacacattttatgcgactggttgacttaaacaacgtcgttacatac 11101 to 11175

gttgcatatgacgagacacaaaaacccatccctgaaaaaagccttgcttcttggcgatgcaatttcagatttacca base pairs
caacgtatactgctctgtgtttttggtagggacttttttcggaacgaagaaccgctacgttaaatggt 11176 to 11250

                                                    MseI          PstI
aaggcaagtgttctgtcaagttcatgcatttctcagtgagcatgctatttaactcttctctgcaggttcaaaaatc base pairs
ttccgttcacaagacagttcaagtacgtaaagagtcactcgtacgataaattgagaagagacgtccaagttttag 11251 to 11325

                                                EcoRI     HhaI   TaqI
accagcctaacgatgtgatggagtatggtggttccccaagaccgaattccagcgctacattcgactcagtcgtaa base pairs
tggtcggattgctacactaccttcataccaccaaggggttctggcttaaggtcgcgatgtaagctgagtcagcatt 11326 to 11400
                                                          HinfI

                                        HaeIII                  MseI
aggtaaaaaaccccgtgaactactactggttggccttcactacgaatatgttaggatttaatttcagaagaaccg base pairs
tccatttttttggggcacttgatgatgaccaaccggaagtgatgcttatacaatcctaaattaaagtcttcttggc 11401 to 11475

                                        PstI                                    AvaII
cctttttttttcttggtgcttcggtactactgcagcaagctcactcttattatcatgtcagacatgttggattggt base pairs
ggaaaaaaaagaaccacgaagccatgatgacgtcgttcgagtgagaataatagtacagtctgtacaacctaacca 11476 to 11550
```

FIG. 5 Continued

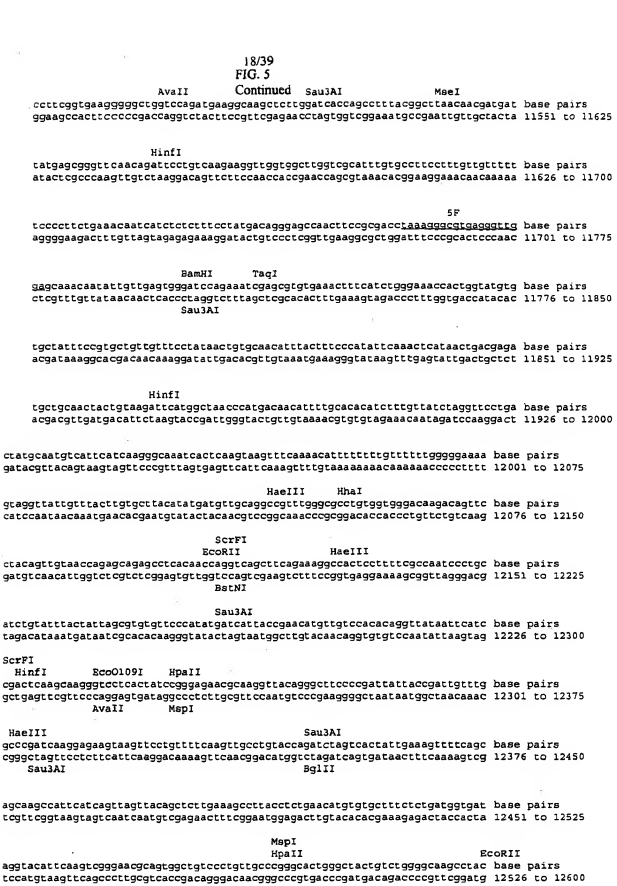


FIG. 5
Continued

```
ScrFI
      HinfI                              PvuII
ctgggtgaatctgaggggagtgaccctctgtaccagctgcctccaagtttcacctctgttggaggacgcactgcg base pairs
gacccacttagactccccctcactgggagacatggtcgacggaggttcaaagtggagacaacctcctgcgtgacgc 12601 to 12675
 BstNI

          Eco0109I                  PstI                          SauJAI
gggcaggcgagggcctcttcctgttggcacccctgcaggggaggtagttgagcagtaaaaggatgacagactcga base pairs
cccgtccgctcccggagaaggacaaccgtggggacgtcccctccatcaactcgtcattttcctactgtctagact 12676 to 12750
           HaeIII                                              1R     BglII
                                     TaqI
gctgagctgggcaacatccagcggcaggagcatttctggttcggttcgattcgggctcacga  base pairs
cgactcgacccgttgtaggtcgccgtcctcgtaaagaccaagccaagctaagcccgagtgct  12751 to 12812
                                     HinfI
```

FIG. 6

PROCESS	WORLD WIDE WEB SITE
sequence format conversion	http://dot.imgen.bcm.tmc.edu:9331/seq-util/Options/readseq.html
reverse complementation	http://dot.imgen.bcm.tmc.edu:9331/seq-util/Options/revcomp.html
sequence translation	http://dot.imgen.bcm.tmc.edu:9331/seq-util/Options/sixframe.html
protein information	http://www.expasy.ch/tools
sequence alignments using Clustal W	http://dot.imgen.bcm.tmc.edu:9331/multi-align/Options/clustalw.html
sequence comparisons using BLAST 2.0	http://www.ncbi.nlm.nih.gov/gorf/bl2.html
sequence searches using BLAST 2.0	http://www.ncbi.nlm.nih.gov/blast/blast.cgi?Jform=0

21/39

FIG. 7

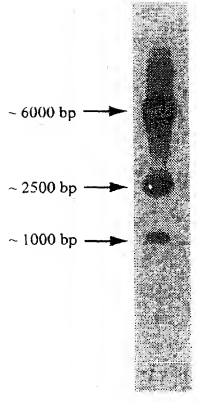


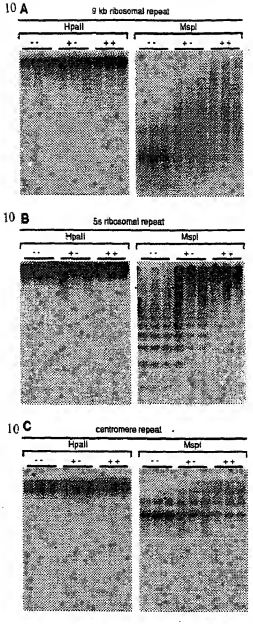
FIG. 8

```
                                                                              NLS
zmet2    1 mapsspspaaptrvsgrkraakaeeihqnkeeeeevaaassakrsrkaassgkkpksppkqakpgrkkkgdaemkepveddvcaeepdeeelaMGeeeaeeeqamQeEvvavaAGsPg
CMT1     1 -------------------------------------------------------------------------------------------MAarnkqkkraEpEsdlcfAGkP-

            NLS
zmet2  118 kkrvgrrnAaaaAgdHePefigSpvaadEArSnwPkryGrstaAKKpDEEeelkARCHYrsAkVDnVVycLgDDVYVkAgeneadYIGRItEfFEGtDqchYftcRWFFRaEDTVIn
CMT1    23 -------mSvveStiRwPhryqSkktklQApTkkPankG----GKKeDEEiikqAKCHFdkAlVDgVLinLnDDVYVtGlpgklkFIAKViElFEAdDgvpYcrfRWYYRpEDTLIe

                                                                                                                 Motif I
zmet2  235 slvsisvdgEkhdPRRVFLSeEkNDNvLdCIiSKVkIvHVd-PnMdpKakaqLIesCDlYYDMsYsVaYstFaniSsENgqsGSDtASgISSDdvdLetsssMptrtatLLDLYSGC
CMT1   130 rfs------ElvqPKRVFLSnDeNDNpLtCIwSKVnIaKVplPkItsRieqrVIppCDyYYDMkYeVpYlnFt--SaDD---GSDaSSsLSSD-saLncfenLhkdekfLLDLYSGC
zmet1 1084                                                                                                          LDIFAGC
MET1  1096                                                                                                          LDIFAGC

       Motif I
zmet2  351 GGMSTGLCLGAALSGLKLeTRWAVDFNsFACQSLKyNHPQTEVRNEkADEFLALLKEWavLCkKY-----------VqDVDsnLasseDQaDEdSp---LDkdEFVVEKlVGIcYGg
CMT1   228 GAMSTGfCMGASISGVKLiTKWSVDiNkFACdSLKlNHPETEVRNEaAEDFLALLKEWkrLCeKFslvsstepvesIsELEdeeveenDdiDEaStgaeLEpgEFeVEKfLGImFGd
zmet1 1091 GGLSeGLqqAGvsf-----TKWAIEYeepAgEAfnkNHPEavV---fvDncnVILKai-------------------------MdkcgDtdDcvSt-------------------se
MET1  1103 GGLShGLkkAGvsd-----AKWAIEYeepAgQAfKqNHPEStV---fvDncnVILRai-------------------------MekggDQdDcvSt-------------------te

                   Chromodomain                                        Motif IV                                 Motif VI
zmet2  454 SD--rEngIyfkVqWeGYgpeeDTWEPiDnLsDCpQKIREFVqEGhKrKILPLPGDVDvICGGPPCQGISGFNRYRNrDePLkDEKNkQMVtFMDIVaYLKPKYVLMENVVDYLKFA
CMT1   352 pqgtgEktLqlmVrWkGYnsayDTWEPysgLgNCkEKLKEYVIDGfKsHLLPLPGtVytVCGGPPCQGISGYNRYRNnEaPLeDQKNqQLLvFLDIIdFLKPnYVLMENVVDLLRFS
zmet1 1156 Aa--------------------------E----qaaKLpEvnin-----nLPVPGEVEfInGGPPCQGfSGmNRFn--qsPwskvqcemILaFLsfaeYfRPRFfLLENVrNfVsFn
MET1  1168 AN--------------------------E----laaKLtEeqks-----tLPLPGqVDfInGGPPCQGfSGmNRFn--qsswskvqcemILaFLsfadYfRPRYfLLENVrtfVsFn

                          Motif VIII
zmet2  569 dGYLGKyALScLVAMkYQaRLGMMvAGcYGLPQfRmRVFLWGAIsSmVLPkYPLPTyDVvVRggaPnaFsqcmVAydETQkp-SLkKALlLGDAISDLPKVqNhqpNDVMEYG-GSP
CMT1   469 kGFLARhAVAsfVAMnYQtRLGMMaAGsYGLPQlRnRVFLWAAqpSekLPpYPLPThEVakKfntPkeFkdlqVGriQmEfl-kLdnALtLADAISDLPpVtNyvaNDVMDYndAAP
zmet1 1236 kGqtfRlAVAsLLeMgYQvRfGILeAGaFGVaQsRkRaFIWAAapgeMLPdWPePmhvfaspelkitlpdgqyyAaarStaggApfRAItVrDtIgDLPKVgNgaskltLEYG-GeP
MET1  1248 kGqtfqltLAsLLeMgYQvRfGILeAGaYGVsQsRkRaFIWAAapeeVLPeWPePmhvfgVpklkislsqglhyAavrStalgApfRpItVrDtIgDLPsVeNgdsrtnkEYk-eva

zmet2  684 KTEFQRYIRLSRkDmLdwSfGegAgPdegKLlDHQPLrLNNDDyERVQqIPvKKGANFRDLkGVRVgaNNivEwDPeiervKLsSGKpLVPDYAMSFIKGKSlKPFGRLWWDEtVPT
CMT1   585 KTEFEnFIsLkRsEtLlpAcG--GdP-trRLfDHQPLvLgDDDlERVsyIPkqKGANYRDMpGVlV-hNNkaEiNPrfr-aKLkSGKnVVPaYAISFIKGKSkKPFGRLWgDEiVnT
zmet1 1352 vSwFQKkIRgS---------------mmvLnDHiskeMNelNliRcQhIPkRpGcDWHDLpdeKVklsNgqmaDl---------ipwcLPNtAkrhnqwKg--lYGRLdWEgnfPT
MET1  1364 vSwFQKeIRgn---------------tiaLtDHickaMNelNliRcklIPtRpGADWHDLpkrKVtlsDgrveEm---------ipfcLPNtAerhngwKg--lYGRLdWqgnfPT

                                    Mu
                          Motif IX  ▽                     Motif X
zmet2  801 VVTRAEPHN--QVIIHPtQaRVLTIRENARLQGFPDyYRLfGpIKEKYIQVGNAVAVPVArALGYcLGQAylGeSEGSDPLyqLPpsFtsvggrtagQARaspvgtpagevveq
CMT1   697 VVTRAEPHN--QcVIHPmQnRVLSVRENARLQGFPDcYKLcGtIKEKYIQVGNAVAVPVGvALGYafGmAsqQlTD-dEPVlkLPfkYp-----ecmQAKdqi
zmet1 1442 sVTdpQPmgkvgMcfHPdQdRIITVREcARsQGFPDsYefaGnIqnKhrQIGNAVppPLAyALGrkLkEAvdkrqEASagV-paP
MET1  1454 sVTdpQPmgkvgMcfHPeQhRILTVREcARsQGFPDsYefaGnInhKhrQIGNAVppPLAfALGrkLkEAlhlkks---P-qhqP
```

FIG. 9

	SAM binding		Cytosine binding	
Motif	M.*Hha*I	*zmet2a*	M.*Hha*I	*zmet2a*
I	Phe18	Try347		
II	Glu40	Gln407		
	Trp41	Trp408		
III	Asp60	Asp428		
IV	Pro80	Pro516	Phe79	Pro515
	Gln82	Gln82	Cys81	Cys517
V	Leu100	Val542		
VI			Glu119	Glu559
			Asn120	Asn560
			Val121	Val561
VIII			Arg165	Arg605
X	Asn304	Asn851		

FIG. 10



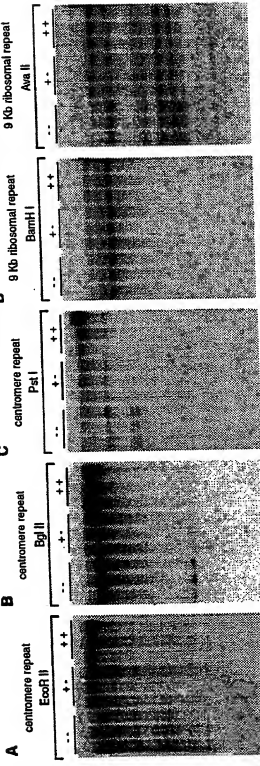
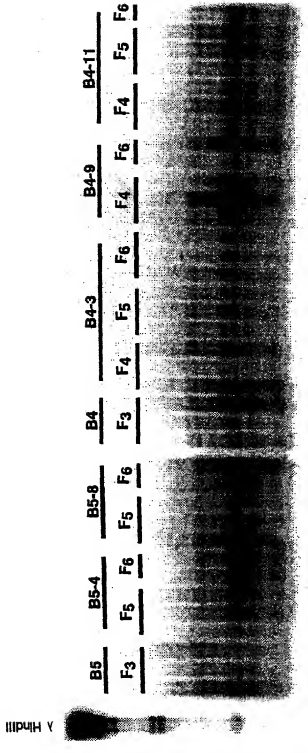

A
centromere repeat
EcoR II
B
centromere repeat
Bgl II
C
centromere repeat
Pst I
D
9 Kb ribosomal repeat
BamH I
E
9 Kb ribosomal repeat
Ava II

FIG. 11

FIG. 12

GENOTYPE	NUMBER OF PLANTS	TOTAL 5mCytosine (%)	% WT levels	% decrease
wild type	3	34.40 ± 0.55	100	0.0
heterozygous zmet2a-mu1	7	32.00 ± 0.90	93.0	7.0
homozygous zmet2a-mu1	5	30.40 ± 0.19	88.4	11.6

FIG. 13


λ HindIII
B5
F3
B5-4
F5
F6
B5-8
F5
F6
B4
F3
B4-3
F4
F5
F6
B4-9
F4
F6
B4-11
F4
F5
F6

28/39

FIG. 14

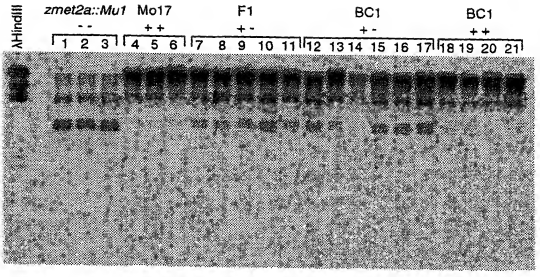


FIG. 15

15 A 15 B 15C

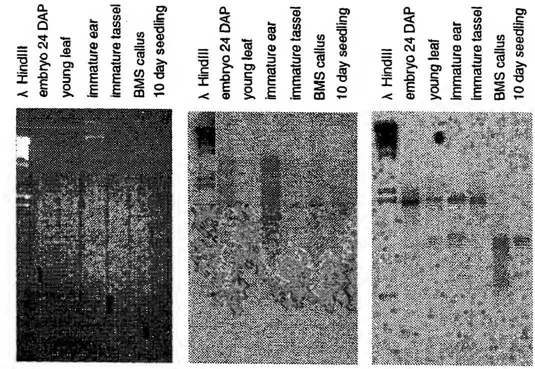


FIG.16

5' LTR

catgc**TGT**TGGGCCATGTGTCTAGTGTTGGCCCATTAACGTGTACA
CATATACTAGAAGTGTGTGTGGTGTAGAGAGAGTGCTGTATGTTTT
CCACATTCCAGAAAAATCC**ACA**TGGTATCAGAGCCAGG
PBS

3' LTR

PPT
GAGGGGGAG**TGT**TGGGCCATGTGTCTAGTGTTGGCCCATTAACGTG
TACACATATACTAGAAGTGTGTGTGGTGTAGAGAGAGTGCTGTATG
TTTTCCACATTCCAGAAAAATCC**ACA**catgc

FIG. 17

Gag **Protease**

```
SPRITE-1  - CYNCGNVGHIARNC    TQVTQLKWILDSGASKH
hopscotch - CQVCSRVGHTALNC    QNGSNVPWYTDTGATDH
retrofit  - CQVCFKRGHTAADC    SYGIDTNWYIDTGATDH
arabpolprt- CSNCGRTGHEKKEC    GKTKLGDIILDSGASHH
copia     - CHHCGREGHIKKDC    SVMDNCGFVLDSGASDH
```

Integrase

```
SPRITE-1  - QVKILRPDN-GTEYVNKGFNAFLSRNGILHQTSCPDTPPQNGVAERKNRHILE
hopscotch - KIIAFQSDW-GGE--YEKLNAHFKTIGIHHQVSCPHTHQQNGAAERKHRHIVE
retrofit  - KIIAMQTDWRGGR--YQKLNSFFAQIGLIIMCHVLTLIRQNGSAERKHRHIVE
arabpolprt- TVKMVRSDN-GTE--FMCLSSYFRENGIIHQTSCVGTPQQNGRVERKHRHILN
copia     - KVVYLYIDN-GREYLSNEMRQFCVKKGISYHLTVPHTPQLNGVSERMIRTITE
```

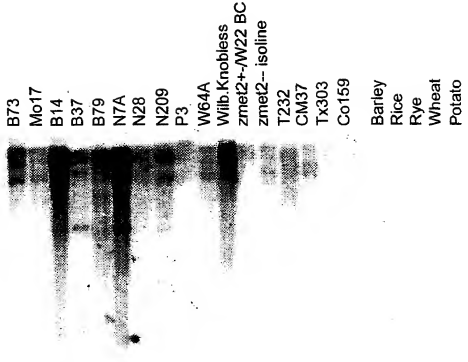
Reverse Transcriptase

```
SPRITE-1  - RYKARLVARGYSQTYGIDYDETFAPVARMSTVRTLISCAANFGMPLYQLDVKNAFLHGDLQEEVYMEIPPG(59)AILAVYVDDIII
hopscotch - RLKARLVAKGFKQQYGIDYDDTFSPVVKHSTIRLVLSLAVSQKWSLRQLDVQNAFLHGILEETVYMKQPPG(59)IYILVYVDDIII
Retrofit  - RYKARLVAKGFKQRYGIDYEDTFSPVVKAATIRILISIAVSRGWSLRQLDVQNAFLHGFLEEEVYMQQPPG(59)MFVLVYVDDIIV
Arabpolprt- RYKARLVVQGNKQVEGEDYKETFAPVVRHTTVRTLLANVAANQWEVYQMDVKNAFLHGDLEEEVYMKLPPG(59)LRVLIYVDDLLI
copia     - RYKARLVARGFTQKYQIDYEETFAPVARISSFRFILSLVIQYNLKVHQMDVKTAFLNGTLKEEIYMRLPQG(59)IYVLLYVDDVVI
```

RNase H

```
SPRITE-1  - DADWGSCLDDRRSTSGYCVFVGG-NLVSWRSKKQSVSRSTAEAEYRAMAIACEMLWIKGLL(25)NPVQHSRTKHVEIDRFF
hopscotch - DADWAGCPDDRKSTGGYALFLGP-NLISWHSKKQSTVSRSSTEAEYKAMAAATAEVIWLQSLL(25)KPIFNAATKHIEVDFHF
retrofit  - DADWAGSIDDRKSTGGFAVFLGS-NLVSWSAKKQSTVSRSSTEAEYKAVAHTTAELIWVQTLL(25)NPVFNAATKHIEVDYHF
arabpolprt- DSDWQSCPLTRRSISAYVVLGG-SPISWKTKKQSTVSHSSAEAEYRAMSYALKEIKWLRKLL(25)NPVFHEATKHIESDCHS
copia     - DSDWAGSEIDRKSTTGYLFMMFDFMLICWMTKRQNSVAASSTEAEYMALFEACREALWLKFLL(25)NPSCHKRAKHIDIKYHF
```

FIG. 18


B73
Mo17
B14
B37
B79
N7A
N28
N209
P3
W64A
Wilb.Knobless
zmet2+-/W22 BC
zmet2-- isoline
T232
CM37
Tx303
Co159
Barley
Rice
Rye
Wheat
Potato

33/39

FIG. 19

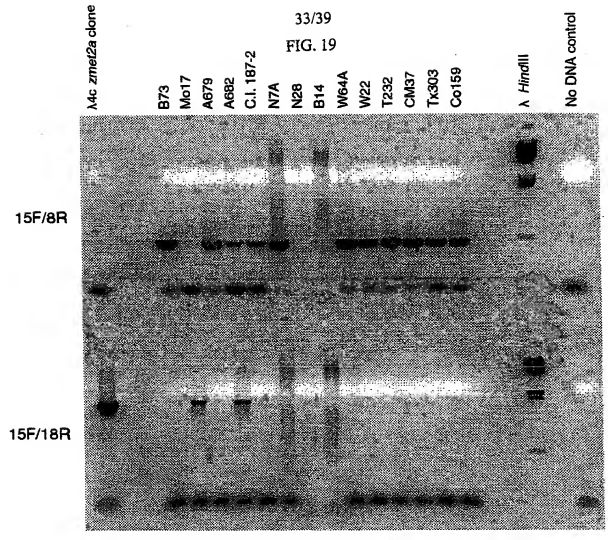


FIG. 20

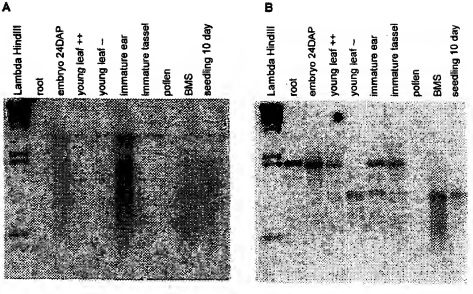

A
Lambda HindIII
root
embryo 24DAP
young leaf ++
young leaf –
immature ear
immature tassel
pollen
BMS
seedling 10 day
B
Lambda HindIII
root
embryo 24DAP
young leaf ++
young leaf –
immature ear
immature tassel
pollen
BMS
seedling 10 day

FIG. 21

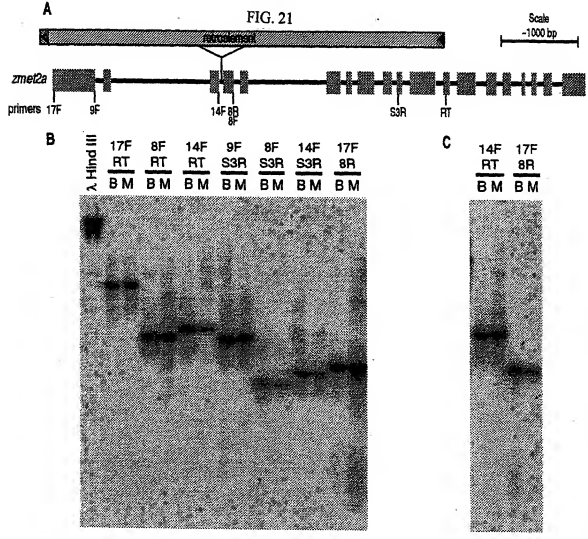

A
Scale
~1000 bp
zmet2a
primers 17F
9F
14F
8R
8F
S3R
RT
B
λ Hind III
17F RT
8F RT
14F RT
9F S3R
8F S3R
14F S3R
17F 8R
B M
C
14F RT
17F 8R

FIG. 22

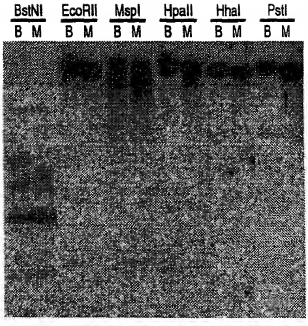


FIG. 23

GGGAATTCGATTACTCACTATAGCGCTCGAGCGGCCGCCCGGGCAGGTTCGAAAACCATC
AACCTAACGATGTAATGGAGTATGGTGGTTCCCCCAAGACAGAGTTCCAGCGCTACATTC
GACTTGGTCGTAAAGACATGTTGGATTGGTCGTTTGGTGAGGAGGCTGGTCCAGATGAAG
GCAAGCTCTTGGATCACCAGCCCTTACGGCTTAACAATGATGATTATGAGCGGGTTAAGC
AAATTCCTGTCAAGAAGGGAGCCAACTTCCGTGACCTAAAGGGTGTCAAGGTTGGAGCAA
ATAATGTTGTTGAGTGGGATCCAGAAGTCGAACGTGTGTACCTTTCGTCTGGGAAACCAC
TGGTTCCTGACTATGCGATGTCATTCATCAAGGGCAAATCACTCAAGCCATTCGGGCGCC
TGTGGTGGGACCAGACGGTTcCTACAGTTGTGACCAGAGCAGAGCCTCATAACCAGGTTAT
ATTGCATCCGACTCAAGCAAGAGTCTTGACTATCCGGGAGAACGCAAGGTTACAGGGCTT
CCCCGATTACTACCGATTGTTTGGACCGATCAAGGAGAAGTATATTCAAGTCGGGAACGC
AGTGGCAGTCCCTGTTGCACGGGCACTGGGCTACTGTCTGGGTCAAGCCTACCTGGGTGA
ATCTGACGGGAGTCAGCCTCTGTACCAGCTGCCTGCAAGTTTTACCTCTGTGGGGCGAAC
CGCGGTTCAGGCGAATGCCGCTTCTGTTGGCACTCCTGCGGGGGAGGTAGTCGAGCAGTA
AAAGGATAGCGGAGCAACCCTGGTTGGTATTTTGATTCGAGCCCATCCAGTAGCATGTTT
ACCAATAAATAATCATTGGTCGTGCTGATTCTTATGGTTGGAGATGAATGTATGTAGGGT
GTACTCGAGCTCGAGTGCTTGTTGTACTGTAGGTTGAGGTTTCTCATCCATTGGCCTGCC
TATTTGTGGATGACGTTTCATTTCAGATTAGCAATGTGCTTATTTAAGGTTTCGTCATGT
ACCTGTATTCTACAATCCACTATTGTTTCCAAAGACAGCATTTGATCCTTAAAAAAAACT
GTAAAAAAAAAAAAACAGTGCCCGAAAAGCCGCAAAAAAAAAAAAAAAAAAAACCTGCCC
GGGCGGCCGCTCGAGCCCTATAGTGAGTAATCGAATTCCC

38/39

FIG. 24

EFDYSL*RSSGRPGRFENHQPNDVMEYGGSPKTEFQRYIRLGRKDMLDWS
FGEEAGPDEGKLLDHQPLRLNNDDYERVKQIPVKKGANFRDLKGVKVGAN
NVVEWDPEVERVYLSSGKPLVPDYAMSFIKGKSLKPFGRLWWDQTVPTVV
TRAEPHNQVILHPTQARVLTIRENARLQGFPDYYRLFGPIKEKYIQVGNA
VAVPVARALGYCLGQAYLGESDGSQPLYQLPASFTSVGRTAVQANAASVG
TPAGEVVEQ*

FIG. 25

667 KVQNHQPNDVMEYGGSPKTEFQRYIRLSRKDMLDWSFGEGAGPDEGKLLDHQPLRLNNDD 726
+ +NHQPNDVMEYGGSPKTEFQRYIRL RKDMLDWSFGE AGPDEGKLLDHQPLRLNNDD
15 RFENHQPNDVMEYGGSPKTEFQRYIRLGRKDMLDWSFGEEAGPDEGKLLDHQPLRLNNDD 74

727 YERVQQIPVKKGANFRDLKGVRVGANNIVEWDPEIERVKLSSGKPLVPDYAMSFIKGKSL 786
YERV+QIPVKKGANFRDLKGV+VGANN+VEWDPE+ERV LSSGKPLVPDYAMSFIKGKSL
75 YERVKQIPVKKGANFRDLKGVKVGANNVVEWDPEVERVYLSSGKPLVPDYAMSFIKGKSL 134

787 KPFGRLWWDETVPTVVTRAEPHNQVIIHPTQARVLTIRENARLQGFPDYYRLFGPIKEKY 846
KPFGRLWWD+TVPTVVTRAEPHNQVI+HPTQARVLTIRENARLQGFPDYYRLFGPIKEKY
135 KPFGRLWWDQTVPTVVTRAEPHNQVILHPTQARVLTIRENARLQGFPDYYRLFGPIKEKY 194

847 IQVGNAVAVPVARALGYCLGQAYLGESEGSDPLYQLPPSFTSVGGRTAGQARASPVGTPA 906
IQVGNAVAVPVARALGYCLGQAYLGES+GS PLYQLP SFTSV GRTA QA A+ VGTPA
195 IQVGNAVAVPVARALGYCLGQAYLGESDGSQPLYQLPASFTSV-GRTAVQANAASVGTPA 253

907 GEVVEQ 912
GEVVEQ
254 GEVVEQ 259